DOCUMENT RESUME

ED 166 257 TM 008 333

AUTHOR	Rudner, Lawrence M.
TITLE	A Short and Simple Introduction to Tailored Testing.
PUB DATE	Mar 78
NOTE	15p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Williamsburg, Virginia, March, 1978)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Ability; Bayesian Statistics; *Individual Tests; *Item Analysis; *Item Banks; *Mathematical Models; Test Construction; *Testing; Testing Problems; Test Items
IDENTIFIERS	Adaptive Testing; *Computer Assisted Testing; Fixed Steps Procedure; Flexilevel Tests; Robbins Monro Process; *Tailored Testing; Test Length

ABSTRACT

Tailored testing provides the same information as group-administered standardized tests, but can do so using fewer items because the items administered are selected for the ability of the individual student. Thus, tailored testing offers several advantages over traditional methods. Because individual tailored tests are not timed, anxiety is reduced and examinee motivation is improved. Economic advantages involve reduced test time, immediate availability of results, and reduced personnel requirements. Effective tailoring occurs at the item level, involving two steps: estimation of the examinee's ability from his or her previous responses, and selection from an item bank of the item likely to measure most effectively. Five methods of estimating item difficulty or appropriateness are: (1) Robbins Monro procedure; (2) fixed step size; (3) flexilevel, which requires a smaller item pool; (4) Bayesian procedures; and (5) stratified-adaptive or stradaptive procedures. (Author/GDC)

A Short and Simple Introduction to Tailored Testing

Lawrence M. Rudner

Gallaudet College
The Model Secondary School for the Deaf
Kendall Green
Washington, D.C. 20002

Paper presented at the Eastern Educational Research Association Annual Meeting, Williamsburg, VA, March, 1978.

This paper was sponsored in part with funds from P.L. 89-694. However, the opinions and views expressed herein are those of the author and do not necessarily reflect those of the Model Secondary School for the Deaf or the U.S. Department of Health, Education and Welfare.

A Short and Simple Introduction to Tailored Testing

Lawrence M. Rudner

As probablistic models with sample independent item description(s) latent trait theories have several appealing features: (1) the performance of an examinee of known ability on a given calibrated item can be predicted; (2) items which were calibrated on different populations can readily be combined to form an item pool of predictable characteristics; and (3) item descriptions are independent of each other. These and other features, along with the advent of readily accessible high speed computers have spawned a re-examination of test development and test usage procedures. In this symposium, Drs. Robertson, Rentz and Durovic have already described how a latent trait model can be applied to the practical issues of test development, equating, and item bias.

Prior to the popularization of group testing procedures, most tests were individually administered and tailored to examinees. At the present time some of the largest developers and users of educational tests, i.e. the United States Department of Defense and the United States Civil Service Commission are re-examining this idea in light of the theoretical and practical benefits of latent trait models.

The term "tailored testing" serves as a generic for any procedure by which particular items or groups of items are selected and administered to any individual examinee based on an estimate of his or her ability. For example, an examinee's grade placement can serve as an initial estimate of an examinee's ability and used to route an examinee to a set of items, e.g. a particular level of an achievement test battery. Such a procedure fits the definition in the broadest sense at the word. However, the procedure will misroute large numbers of both very able and lesser able students, and is not considered in most discussions of tailored testing.

Effective tailoring occurs at the item level; effective in the sense that examinees across a wide continuum of ability are administered items appropriate to their competence and are therefore not misrouted. The task involves two basic iterated steps:

1. Estimation of the examinee's ability from his or her previous responses.
2. Selection of the item likely to measure most effectively at the presently estimated ability level (Lord, 1977).

Prior to tailoring, an item pool must be developed and item characteristics computed. Since the best items will be those whose difficulties most closely match the examinee's ability, all item-level tailoring schemes use an index of item difficulty (either the proportion of examinees responding correctly, the Rasch model easiness parameter, or the 2 and 3 parameter model location or difficulty parameter). Some schemes will also incorporate item discrimination indices and/or guessing indices.

The interested reader is referred to Angoff and Huddleston (1958), Ferguson (1969), Krathwohl and Huyser (1956), Linn, Rock and Cleary (1968, 1972), Lord (1971a, 1971b), Olivier (1973), Owen (1969), Vale (1975) and Weiss (1973, 1974) for further descriptions of the presented approaches and others and to Cleary, Linn and Rock (1968), Linn, Rock and Cleary (1968, 1972), Lord

(1971a), Urry (1977), Vale (1975), and Vale and Weiss (1975) for some evaluations.

Robbins Monro Procedures

In Robbins Monro procedures, the difficulty of the (i + 1)st item to be administered is determined by the rule

$$b(i + 1) = d_i (M_i - g) + b_i$$

where b_i is the difficulty of the ith administered item,

d_i is a descending sequence of positive numbers,

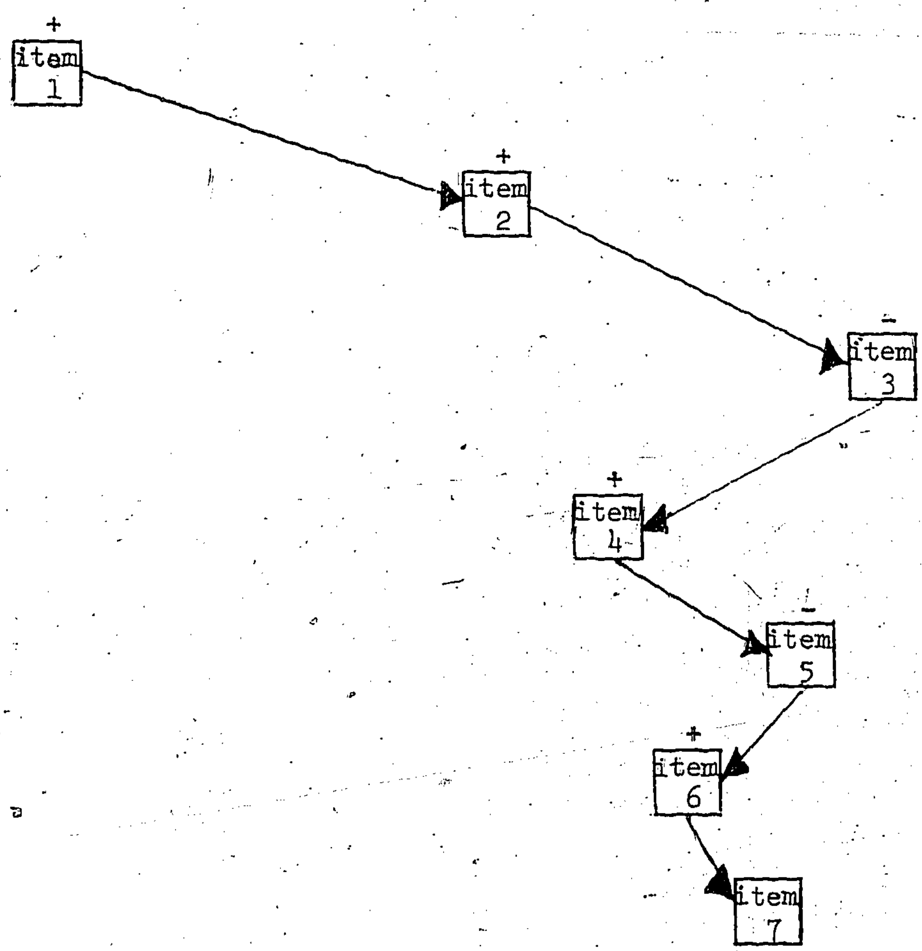
M_i is the response to item i. ($M_i = 1$ when correct, $M_i = 0$ when incorrect)

g is an offset parameter

This procedure is illustrated in Figure 1. The examinee took eight items. Harder items are administered after each correct response, easier after each incorrect response. The difference in item difficulty between consecutively administered items decreases proportionately since d_i is a descending sequence. The process continues converging on the point at which the item difficulty is equal to the examinee's ability and is terminated when a satisfactory estimate is achieved. After n items are administered, the difficulty of the (n + 1)st item can then be used as the estimate of the examinee's ability.

FIGURE 1

A Hypothetical Example of the Robbins-Monro Tailored Testing Procedure



Fixed Step Size

Rather than using the decreasing step size governed by d_i in the Robbins Monro procedures, d_i can be held constant and the (i + 1)st item to be administered can be selected by

$$b_{(i+1)} = b_i + d\ (M_i + g)$$

Figure 2 illustrates this procedure. This procedure can never truly converge on the point where the item difficulty equals the examinee's ability. The difficulty of the administered items will vascillate between being just above the examinee's ability and just below it. The average difficulty of the administered item can be used as an estimate of the examinee's ability.

Flexilevel

One practical limitation with the Robbins Monro and Fixed Step Size procedures is the need for extremely large item pools. In theory, the later procedure will require n(n + 1)/2 items, the former substantially more.

The flexilevel procedure routes the examinee to the next less difficult unadministered item following an incorrect response. Following a correct response, the examinee is routed to the next more difficult unadministered item. Thus, the difficulty of the (i + 1)th item is based on the available item pool. The procedure is illustrated in Figure 3. After the item whose difficulty most closely matches the examinee's ability is administered, the selected item oscillates between being substantially too easy and substantially too difficult for the examinee.

FIGURE 2

A Hypothetical Example of the

Fixed Steps Tailored Testing Procedure

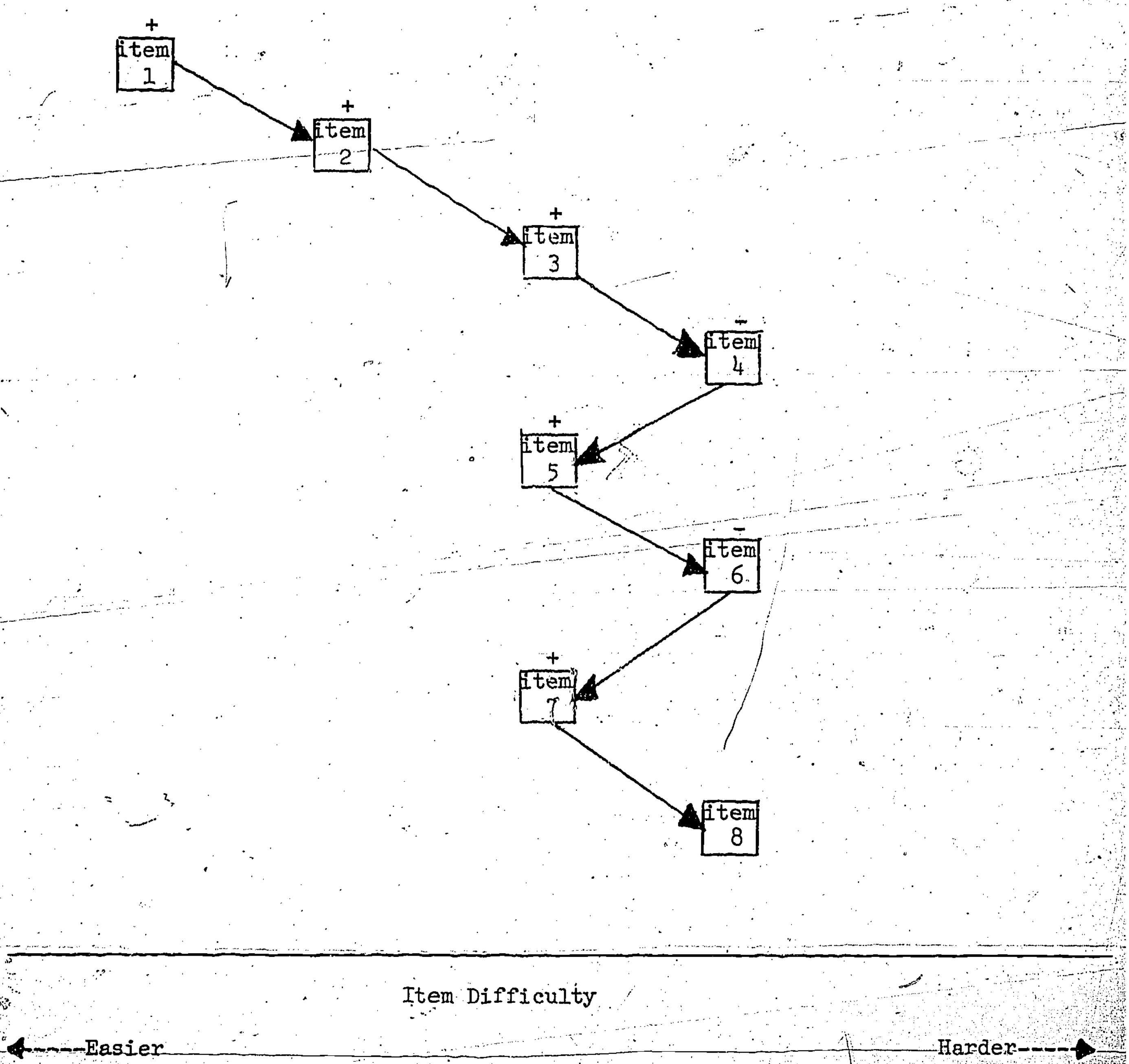
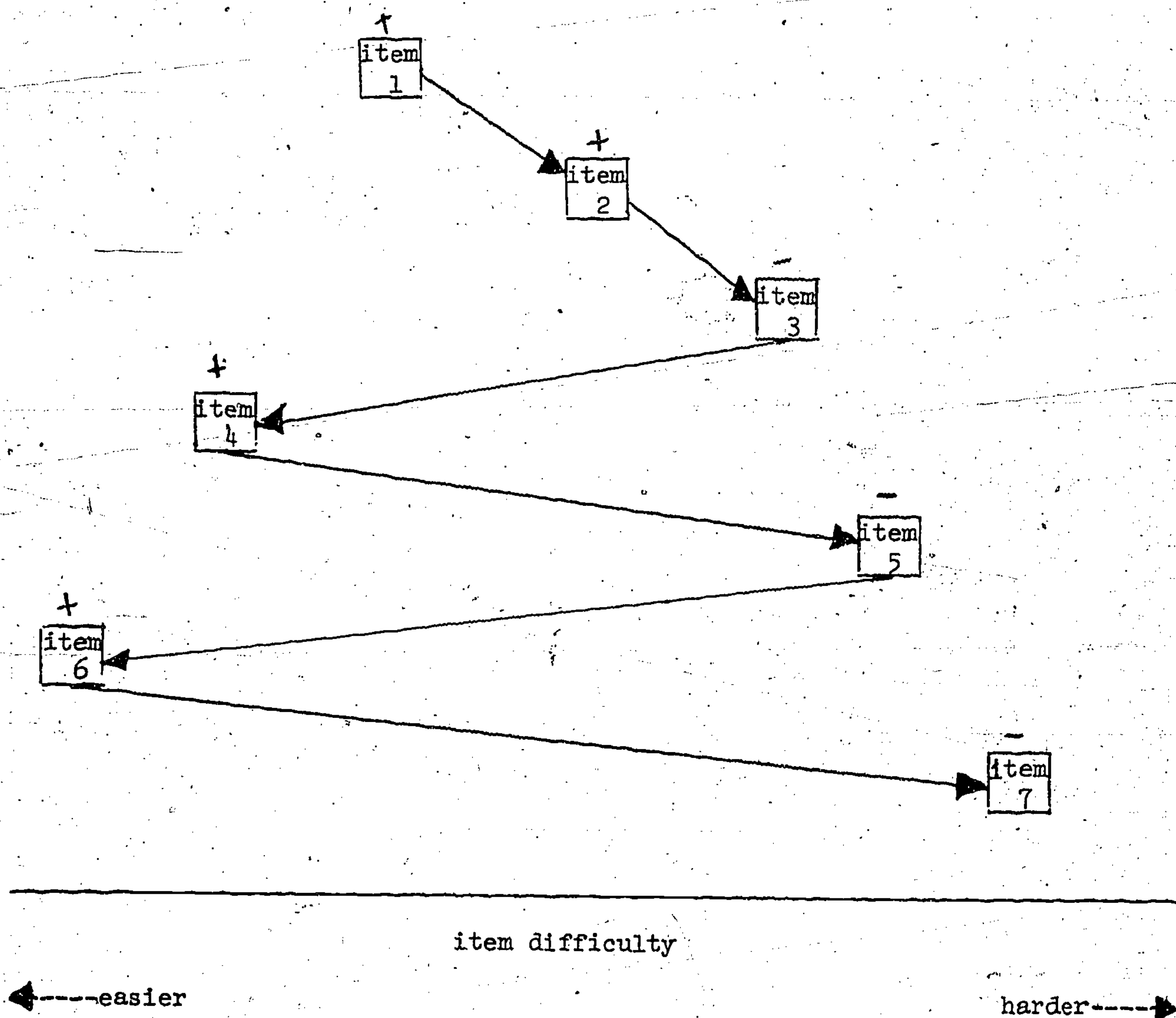


FIGURE 3

A Hypothetical Example of the

Flexilevel Tailored Testing Procedure



Bayesian Procedures

Bayes theorem can be written as

$$P(A|B) = K \cdot P(B|A) \cdot P(A)$$

where K is a constant

$P(A|B)$ denotes the probability of A given B

Substituting estimated ability for A and item response for B, the theorem is well suited to a measurement model which specifies the probability of a correct response given an examinee's ability $P(B|A)$. Assuming a normal prior, $P(A)$ is normal, and using a latent trait item response model, an estimate of examinee ability can be inferred from each item response. After each item is administered, the obtained posterior estimate of ability serves as the prior estimate for the next item. Items are selected so as to minimize a loss function. When guessing is assumed to be void and item discriminations approximately equal (as with the Rasch model), items are selected such that the difference between item difficulty and the estimated ability is the minimum possible within the item pool restraints. When guessing is a factor, the optimal difficulty is a bit less than the examinee's estimated ability. Testing is terminated when the standard error of estimation of ability is sufficiently small or when a maximum number of items have been administered.

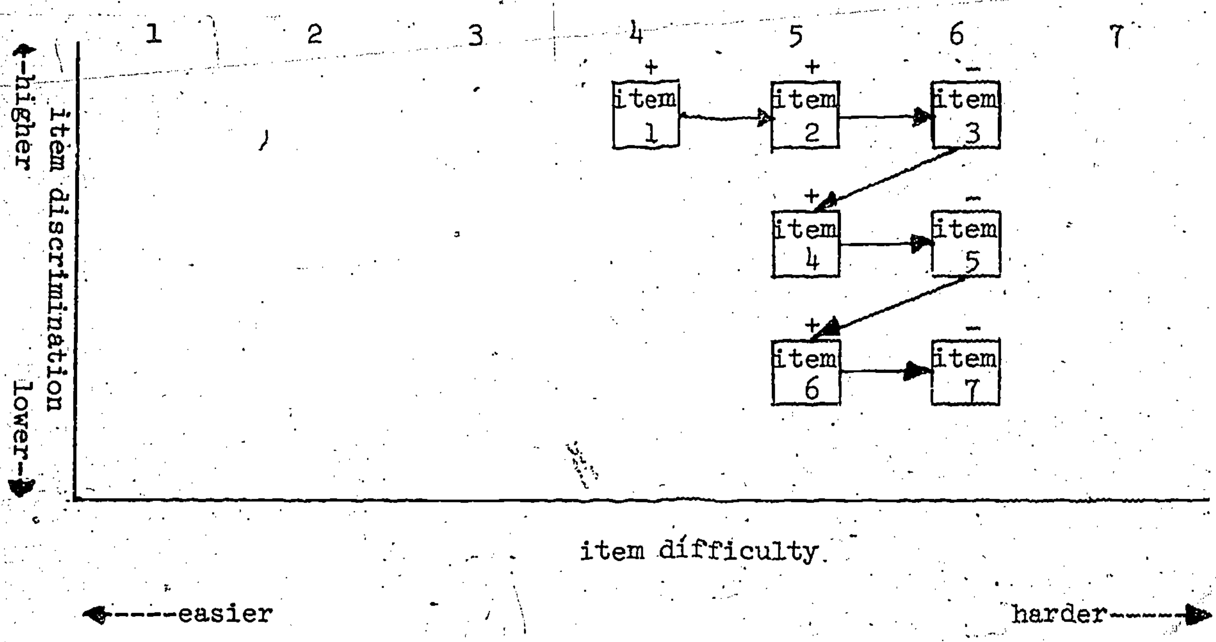
Stratified-adaptive (stradaptive) Procedures

Two of the main advantages of the Bayesian procedures are that prior, non-test estimates of ability can be used to allow for multiple entry points and that item information in addition to item difficulty can be used. The main limitation is that before an item is selected to be administered, all unadministered

FIGURE 4

A Hypothetical Example of

Stratified-Adaptive Tailored Testing



items must be evaluated for their effectiveness. The stratified-adaptive procedures seek to circumvent this limitation while retaining the benefits.

Figure 4 illustrates stratified-adaptive tailored testing. Items are arranged into strata of increasing difficulty. In this example, seven strata are defined. Within each stratum, the items are arranged in order of their discriminability. Testing begins by administering the best discriminating item in the stratum whose difficulty level most closely matches the prior estimate of the examinee's ability, or in the median strata when no prior is available. After a correct response, the procedure routes the examinee to the best discriminating unadministered item in the next more difficult strata. Following each incorrect response, the examinee is administered the best discriminating unadministered item in the next less difficult strata.

Conclusions

One can safely say that the majority of standardized tests are administered to groups of students rather than are individually administered, use a paper and pencil format with separate question and response sheets, present all items to all examinees in a fixed order, and have set time limits. Tailored testing seeks to provide the same information as such group tests, but by presenting fewer items. Since the items are tailored to the students ability and since fewer items are administered, tailored tests can offer several advantages over traditional assessment instruments.

In a comprehensive 1973 review of the advantages and limitations of tailored testing, Weiss and Betz point out that the high degree of standardization of group tests introduces problems of time limits, answer sheets, test compromise, administrator variables, and item arrangements as they affect whole groups and as they affect certain subgroups of individuals. Individual tests, on the other hand, are untimed - thus minimumizing anxiety and increasing accuracy; better maintain motivation - since guessing and fatigue is reduced.

An additional factor which is not often discussed is the managability of tailored testing. Test results are immediately available, testing on demand is possible, trained test administrators are not required, and examinee time is reduced. As a consequence, tailored tests can offer economic as well as psychometric advantages over conventional tests.

REFERENCES

Angoff, Witt, and Huddleston, E.M. The multi-level experiment: A study of a two-level test system for the College Board Aptitude Test. Princeton, N.J.: Educational Testing Service, 1958.

Cleary, T.A., Linn, R.L. and Rock, D.A. An exploratory study of programmed tests. Educational and Psychological Measurement, 1968, 28, 345-360.

Ferguson, R.L. The development, implementation and evaluation of a computer assisted branched test for a program of individually prescribed instruction, Unpublished doctoral disseration. University of Pittsburg, 1969.

Krathwohl, D.R. and Huyser, R.J. The sequential item test (SIT), American Psychologist, 1956, 2, 419.

Linn, R.L., Rock, P.A. and Cleary, T.A. The development and evaluation of several testing methods, Educational and Psychological Measurement, 1968, 29, 129-146.

Linn, R.L., Rock, D.A. and Cleary, T.A. Sequential testing for dichotomous decisions. Educational and Psychological Measurement, 1972, 32, 85-95.

Lord, F.M. Robbins-Monro procedures for tailored testing Educational and Psychological Measurement, 1971a, 31, 3-31.

Lord, F.M. The self-scoring flexilevel test. Journal of Educational Measurement, 1971b, 8, 147-157.

Lord, F.M. Practical Applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14(2), 117-138.

Olivier, P. An overview of tailored testing. Unpublished doctoral preliminary examination paper, Florida State University, Tallahassee, 1973.

Owen, R.J. A Bayesian approach to tailored testing. Princeton, N.J.: Educational Testing Service, 1969.

Urry, V.W. Tailored testing: A sucessful application of latent trait theory. Journal of Educational Measurement, 1977, 14 (2), 181-196.

Vale, C.D. Strategies of branching through an item pool. In D.J. Weiss (Ed) Computerized adative trait measurement: Problems and prospects, Minneapolis: University of Minnesota Psychometric Methods Programs, 1975.

REFERENCES

Vale, C.D. and Weiss, D.J. A simulation study of stradaptive ability testing. Minneapolis: University of Minnesota, Psychometric Methods Program, 1975.

Weiss, D.J. The stratified adaptive computerized ability test. Minneapolis, MN: University of Minnesota Psychometric Methods Program, 1973.

Weiss, D.J. Strategies of adaptive ability measurement, Minneapolis: University of Minnesota, Psychometric Methods Program, 1974.

Weiss, D.J. and Betz Ability measurement: Conventional or adaptive? Minneapolis, MN: University of Minnesota, Psychometric Methods Program, 1973.